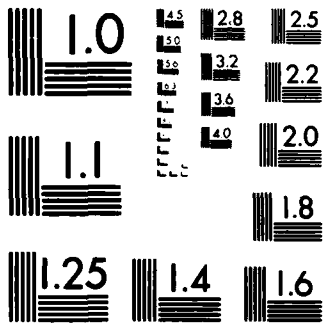
AD-A130 427 A NOTE ON OPTIMAL AND ROBUST SPACING SELECTION(U) 1/1
SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS
R L EUBANK JUN 83 SMU/DS/TR-176 N00014-82-K-0207

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

(13)

SOUTHERN METHODIST UNIVERSITY

A NOTE ON OPTIMAL AND ROBUST SPACING
SELECTION

by

R. L. Eubank

Technial Report No. SMU/DS/TR-176
Department of Statistics ONR Contract

June, 1983

DEPARTMENT OF STATISTICS

DALLAS, TEXAS 75275

A NOTE ON OPTIMAL AND ROBUST SPACING
SELECTION

by

R. L. Eubank

Technial Report No. SMU/DS/TR-176
Department of Statistics ONR Contract

June, 1983

Research sponsored by the Office of Naval Research
Contract N00014-82-k-0207

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

A NOTE ON OPTIMAL AND ROBUST SPACING SELECTION

R. L. Eubank

Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Key Words and Phrases: optimal spacings, robust spacings, location and scale parameters.

ABSTRACT

The problem of quantile selection for the asymptotically best linear unbiased estimators of location and scale parameters is considered. The asymptotic properties of several quantile selection methods for simultaneous parameter estimation are derived and simple approximate solutions are provided. A robust scheme for quantile selection is also developed.

1. INTRODUCTION

Assume that a random sample, $X_1,\ldots,X_n$, has been obtained from a distribution of the form $F(x) = F_0(\frac{x-\mu}{\sigma})$, where F_0 is a known distributional form and μ and σ are, respectively, location and scale parameters. This note is concerned with the estimation of μ and σ by the asymptotically best linear unbiased estimators (ABLUE's) based on $k < n$ sample quantiles.

Define the sample quantile function by

$$\tilde{Q}(u) = X_{(j)}, \quad \frac{j-1}{n} < u \leq \frac{j}{n}, \quad j = 1,\ldots,n, \tag{1.1}$$

where $X_{(j)}$ is the jth sample order statistic; then, given a spacing $T = \{u_1,\ldots,u_k\}$ (k real numbers satisfying $0 < u_1 < \ldots < u_k < 1$) the ABLUE's are easily computed linear functions of the $\tilde{Q}(u_i)$, $i = 1,\ldots,k$. Explicit estimator formulae as well as expressions for the asymptotic efficiency of the ABLUE's relative to the Cramér-Rao lower variance bounds can be found, for example, in Chapter 5 of Sarhan and Greenberg(1962), in Cheng(1975) or Eubank (1981a). Consequently, they will not be repeated here. As these formulae all involve the spacing, T, the problem we address is the selection of spacings that have optimal properties for certain functions of the estimators' asymptotic relative efficiencies (ARE's).

Before proceeding further it should be noted that spacing selection for the ABLUE is related to several other problems including those which derive from i) regression design for time series with Brownian bridge error processes, ii) variable breakpoint $L^2[0,1]$ piecewise constant approximation, iii) grouping selection for the asymptotically most powerful group rank test for two sample location and scale problems and iv) problems of optimal stratification and grouping (see Gastwirth (1966), Adatia and Chan(1981) and Eubank(1982) for discussions of some of the relationships between these problems). Consequently, the results presented here have applications in these areas as well. Of particular importance for this article is the connection between spacing selection and problems i) and ii) which is used, implicitly, in subsequent sections. For more detailed discussions of this relationship and further background material on the ABLUE see Eubank(1981a,b) and Eubank, Smith and Smith(1981).

Let D_k represent the set of all k-element spacings and,

for $T \epsilon D_k$, denote the ABLUE's and their corresponding variance-covariance matrix by $(\mu(T), \sigma(T))^t$ and $\frac{\sigma^2}{n} A(T)^{-1}$, respectively. The joint ARE of the ABLUE's is then given by

$$ARE(\mu(T), \sigma(T)) = |A(T)|/|A| \qquad (1.2)$$

where A is the usual intrinsic accuracy matrix. The construction of spacings which maximize (1.2) is both mathematically and numerically intractable for most distributions. This has lead to consideration of other optimality criteria. For example, Hassanien(1969a, 1969b) and Eisenberger and Posner(1965) choose spacings that minimize the sum of the (asymptotic) estimators' variances which is equivalent to minimizing the trace of $A(T)^{-1}$, denoted tr $A(T)^{-1}$. Another alternative, proposed by Hassanein (1977) is maximization of the sum of the estimators' ARE's or, equivalently, maximizing $trA(T)B^{-1}$ where B is a diagonal matrix consisting of the diagonal elements of A. In Section 2 the asymptotic (as $k \to \infty$) properties of these alternative spacing selection schemes are derived and simple approximate solutions are provided. In each case, the solution is in the form of a density function, h, on [0,1]. The sequence of spacings $\{T_k\}$, $T_k \epsilon D_k$, whose kth element consists of the (k+1)-tiles of h is called the <u>regular sequence generated by h</u>, denoted RS(h), and for k sufficiently large and optimal h, T_k is the proposed approximate solution.

In Section 3 robust spacing selection is considered. The problem, in this case, is to select a (spacing) density that is optimal relative to a known finite set of probability laws. The resulting solution provides an asymptotic analog of a procedure suggested by Chan and Rhodin(1980).

2. ASYMPTOTICALLY OPTIMAL SPACINGS

Denote the quantile function corresponding to F_0 by Q_0.

Assume that F_0 admits a density, f_0, and, hence, a density-quantile function $d_0(u) = f_0(Q_0(u))$, $0 \leq u \leq 1$. Throughout this section we require that both d_0 and the product $d_0 \cdot Q_0$ be twice continuously differentiable on $[0,1]$ and vanish at the ends of the interval. We also adopt the notation

$$\psi(u) = (d_0''(u), (d_0 \cdot Q_0)''(u))^t. \tag{2.1}$$

Using η to denote any one of the criteria considered in Section 1, i.e., $|A(T)|$, $\mathrm{tr}A(T)^{-1}$ or $\mathrm{tr}A(T)B^{-1}$, we now define and illustrate our concepts of asymptotic optimality for spacing sequences. A more detailed development of these topics and other results in this section can be found in Eubank(1981b). A sequence of spacings $\{T_k\}$ is called asymptotically η1-optimal if

$$\lim_{k\to\infty}[\eta(A) - \sup_{T\in D_k} \eta(A(T))][\eta(A) - \eta(A(T_k))]^{-1} = 1 \tag{2.2}$$

and asymptotically η2-optimal if

$$\lim_{k\to\infty}[\inf_{T\in D_k} \eta(A(T)^{-1}) - \eta(A^{-1})][\eta(A(T_k)^{-1}) - \eta(A^{-1})]^{-1} = 1. \tag{2.3}$$

In the case of the determinant criterion it was shown in Eubank (1981a) that a sequence of optimal spacings for (1.2) satisfies

$$\lim_{k\to\infty} k^2[|A| - \sup_{T\in D_k}|A(T)|] = \frac{1}{12}\{\int_0^1[\psi(u)^t A^{-1}\psi(u)]^{1/3}du\}^3 \equiv \lambda_D^3/12 \tag{2.4}$$

and that the RS generated by $h_D(u) = [\psi(u)^t A^{-1}\psi(u)]^{1/3}/\lambda_D$ is asymptotically η1-optimal. Thus (2.2) has the interpretation that, for $\{T_k^*\}$ RS(h_D), $|A| - |A(T_k^*)|$ converges at the same rate (namely $O(k^{-2})$) with the same asymptotic constant $(\lambda_D^3/12)$ as $|A| - \sup_{T\in D_k}|A(T)|$. Similar interpretations hold for the other cases that are considered.

If η is now taken as the sum of ARE's it follows by arguments similar to those in Eubank (1981a) and Theorem 4.1 of Sacks and Ylvisaker (1968) that a sequence of spacings obtained by maximizing $\mathrm{tr}A(T)B^{-1}$ satisfies

$$\lim_{k\to\infty} k^2[\mathrm{tr}AB^{-1} - \sup_{T\varepsilon D_k} \mathrm{tr}A(T)B^{-1}]$$
$$= \frac{1}{12}\{\int_0^1 [\psi(u)^t B^{-1}\psi(u)]^{1/3} du\}^3 \equiv \lambda_S^3/12. \tag{2.5}$$

An approximate (asymptotic) solution is provided by the RS obtained from $h_S(u) = [\psi(u)^t B^{-1}\psi(u)]^{1/3}/\lambda_S$ which is asymptotically η1-optimal. For spacings chosen to minimize the sum of the estimators variances we have (see Theorem 4.5 of Sacks and Ylvisaker(1968) and the subsequent remark)

$$\lim_{k\to\infty} k^2[\inf_{T\varepsilon D_k} \mathrm{tr}\, A(T)^{-1} - \mathrm{tr}A^{-1}]$$
$$= \frac{1}{12}\{\int_0^1 [\psi(u)^t A^{-2}\psi(u)]^{1/3} du\}^3 \equiv \lambda_V^3/12. \tag{2.6}$$

An asymptotically η2-optimal sequence for this criterion is provided by the RS generated by $h_V(u) = [\psi(u)^t A^{-2}\psi(u)]^{1/3}/\lambda_V$.

The diagonal nature of B has the consequence that h_S will be easier to use, in general, than h_D or h_V for spacing computation. Moreover, we see from (2.4) and (2.5) that maximizing $|A(T)|$ and $\mathrm{tr}A(T)B^{-1}$ are asymptotically equivalent procedures for symmetric distributions as, in this case, $B = A$ (this explains the similarity between these two solutions observed by Kulldorff(1963) for the normal distribution). The same cannot be said for spacings which minimize the sum of the variances.

The computation of asymptotically optimal spacings from h_D, h_S, and h_V will usually require the use of numerical methods (c.f. Eubank(1981a)). A distribution admitting a closed form solution is the Cauchy where asymptotically optimal spacings are provided by uniformly spaced points (i.e., $T_k = \{\frac{i}{k+1};\ i = 1,\dots,k\}$) in all three cases.

3. ROBUST SPACING SELECTION

In this section we relax the assumption that F_0 is known

precisely and assume, instead, that F_0 is known only to belong to a given finite set of probability laws, $\mathcal{L}$. The problem now is to select spacings that are robust relative to $\mathcal{L}$.

Consider first the case of location parameter estimation. For L, $G \epsilon \mathcal{L}$ let $T(G) \epsilon D_k$ denote the optimal spacing for G and let $ARE(\mu(T(G))|L)$ be the ARE for T(G) when L is the true underlying distribtuion. Chan and Rhodin(1980) suggest choosing a spacing $T(G^*)$ that satisfies

$$\min_{L \epsilon \mathcal{L}} ARE(\mu(T(G^*))|L) = \max_{G \epsilon \mathcal{L}} \min_{L \epsilon \mathcal{L}} ARE(\mu(T(G))|L). \tag{3.1}$$

This solution provides a candidate for F_0, namely G^*, and μ is estimated accordingly. A spacing selected using (3.1) is an element of $\{T(G); G \epsilon \mathcal{L}\}$ which maximizes the guaranteed asymptotic relative efficiency (GARE), $\min_{G \epsilon \mathcal{L}} ARE(\mu(T(G))|L)$, and is robust in this sense of providing maximum GARE over the optimal spacings for laws in $\mathcal{L}$. A disadvantage of this approach is that tedious computations must be performed for each value of k. We now present an asymptotic (as $k \to \infty$) alternative to (3.1) that alleviates this difficulty.

Let $\mathcal{H}$ denote a finite set of bounded piecewise continuous density functions on [0,1] where, for $h \epsilon \mathcal{H}$, the set of points where 1/h vanishes or is discontinuous is assumed to have content zero and neither 0, nor 1 as an accumulation point. Also assume that for each $L \epsilon \mathcal{L}$ the corresponding density-quantile function d_L is in $C^2(0,1) \cap L^2[0,1]$ and monotone near 0 and 1. It then follows from Pence and Smith(1981) that for any $h \epsilon \mathcal{H}$ and $L \epsilon \mathcal{L}$ if $\{T_k\}$ is RS(h) then

$$\lim_{k \to \infty} k^2 a_{11}(L)[1-ARE(\mu(T_k)|L)] = \frac{1}{12} \int_0^1 [d_L''(u)]^2 [h(u)]^{-2} du \tag{3.2}$$

where $a_{11}(L)$ is the element that corresponds to μ in the infor-

mation matrix for L, A_L say. Thus, what distinguishes between the performance of spacing sequences generated by the densities in $\mathcal{H}$ is, for fixed L, the asymptotic constant in (3.2). An element of $\mathcal{H}$ that is optimal relative to $\mathcal{L}$ is therefore provided by: Choose $h^* \epsilon \mathcal{H}$ to satisfy

$$\max_{L\epsilon\mathcal{L}} \int_0^1 [d_L''(u)]^2[h^*(u)]^{-2}du = \min_{h\epsilon\mathcal{H}} \max_{L\epsilon\mathcal{L}} \int_0^1 [d_L''(u)]^2[h(u)]^{-2}du. \quad (3.3)$$

One can then estimate μ using, for instance, the law at which $\int_0^1 [d_L''(u)]^2[h^*(u)]^{-2}du$ is minimized. Alternative conditions under which (3.3) is valid can be deduced from Theorem 4.4 of Pence and Smith(1981) and Theorem 3.1 of Sacks and Ylvisaker(1968).

A logical choice for $\mathcal{H}$ would seem to be the set of optimal densities for location parameter estimation corresponding to the various laws in L, $\{|d_L''(u)|^{2/3}/\int_0^1 |d_L''(s)|^{2/3}ds;\ L\epsilon\mathcal{L}\}$ (see Eubank(1981a)). Choosing $\mathcal{H}$ in this manner we now compare spacings selected using (3.3) to those obtained by Chan and Rhodin(1980) using (3.1). As they restrict attention to laws that are (essentially) members of the Tukey lambda family we shall do likewise and, also for comparison purposes, take $k = 5$. Using $L(\lambda)$ to denote that member of the Tukey lambda family having shape parameter λ, a comparison of spacing GARE's for the two procedures is provided in Table 1 for a few selected choices of $\mathcal{L}$. It is important to note that a spacing obtained using (3.1) is not a spacing that maximizes the GARE over all $T\epsilon D_k$ (c.f. Chan and Rhodin(1980, p. 236)). Thus spacings obtained using the asymptotic approach may, in fact, result in larger GARE's as is illustrated by the case of $\mathcal{L} = \{L(-.1), L(0), L(.1), L(.14)\}$ and $\mathcal{L} = \{L(-.6), L(-.5), L(-.4), L(-.3)\}$.

TABLE 1. Comparison of Spacing GARE's

L	Optimal Spacings	Asymptotic Solution
{L(0),L(.1),L(.14)}	.9332	.9111
{L(-.1),L(0),L(.1),L(.14)}	.9015	.9080
{L(-.6) ,L(-.5),L(-.4),L(-.3)}	.9568	.9581
{L(-2.0),L(-1.8),L(-1.6), L(-1.4),L(-1.2)}	.9207	.9056

Of course there is no reason to restrict attention to location parameter estimation. A scale parameter version of (3.3) is readily obtained by replacing d_L with the product $d_L \cdot Q_L$ in the previous discussion where Q_L is the quantile function for $L \varepsilon \mathcal{L}$. For the estimation of both μ and σ observe that if $h \varepsilon \mathcal{H}$ and $\{T_k\}$ is RS(h), then from the preceeding comment and (3.2)

$$\lim_{k\to\infty} k^2 \mathrm{tr}[A_L - A_L(T_k)]B_L^{-1} = \frac{1}{12}\int_0^1 \{\psi_L(u)^t B_L^{-1} \psi_L(u)\}[h(u)]^{-2} du \quad (3.4)$$

where ψ, A(T) and B have now been subscripted to indicate their dependence on L. Thus one approach to simultaneous parameter estimation can be based on (3.4). In this case $\mathcal{H}$ might be chosen to consist of the optimal densities for $\mathrm{tr}A_L(T)B_L^{-1}$, $L \varepsilon \mathcal{L}$, given in Section 2. Analogs of (3.4) can also be obtained for the other criteria that have been considered.

REFERENCES

Adatia, A. and Chan, L. K. (1981). Relations between stratified grouped and selected order statistics samples. Scand. Actuarial J.,193-202.

Chan, L. K. and Rhodin, L. S. (1980). Robust estimation of location using optimality chosen sample quantiles. Technometrics 22, 225-237.

Cheng, S. W. (1975). A unified approach to choosing optimum quantiles for the ABLE's. J. Amer. Statist. Assoc. 70, 155-159.

Eisenberger, I. and Posner, E. C. (1965). Systematic statistics used for data compression in space telemetry. J. Amer. Statist. Assoc. 60, 97-133.

Eubank, R. L. (1981a). A density-quantile function approach to optimal spacing selection. Ann Statist. 9, 494-500.

Eubank, R. L. (1981b). A regression design approach to optimal and robust spacing selection. Tech. Rep. No. 144, Dept. of Statist., Southern Methodist University.

Eubank, R. L. (1982). A quantile domain perspective on the relationships between optimal grouping, spacing and stratification problems. Statistics and Probability Letters 1, 69-73.

Gastwirth, J. L. (1966). On robust procedures. J. Amer. Statist. Assoc. 61, 929-948.

Hassanein, K. M. (1969a). Estimation of the parameters of the extreme value distribution by use of two or three order statistics. Biometrika 56, 429-436.

Hassanein, K. M. (1969b). Estimation of the parameters of the logistic distribution by sample quantiles. Biometrika 56, 684-687.

Hassanein, K. M. (1977). Simultaneous estimation of the location and scale parameter of the gamma distribution by linear functions of order statistics. Scandinavian Actuarial J., 88-93.

Kulldorff, G. (1963). On the optimum spacing of sample quantiles from the normal distribution, Part I. Skand. Aktuarietidskr. 46, 143-156.

Pence, D. D. and Smith, P. W. (1981). Asymptotic properties of best $L_p[0,1]$ approximation by splines. SIAM J. Math. Anal. 13, 409-420.

Sacks, J. and Ylvisaker, D. (1968). Designs for regression problems with correlated errors; many parameters. Ann. Math. Statist. 39, 40-69.

Sarhan, A. E. and Greenberg, B. G. (eds.) (1962). Contributions to Order Statistics. New York: John Wiley.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SMU/DS/TR-176	2. GOVT ACCESSION NO. A130 427	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NOTE ON OPTIMAL AND ROBUST SPACING SELECTION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER 176
7. AUTHOR(s) RANDALL L. EUBANK		8. CONTRACT OR GRANT NUMBER(s) N00014-82-k-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-479
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE June, 1983
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of quantile selection for the asymptotically best linear unbiased estimators of location and scale parameters is considered. The asymptotic properties of several quantile selection methods for simultaneous parameter estimation are derived and simple approximate solutions are provided. A robust scheme for quantile selection is also developed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

8-83

DTIC